

THE IMPACT OF SIRE DAUGHTER VARIABILITY
AND REPEATABILITY ON THE ACCURACY
OF SIRE SELECTION

by

John Springmann Clay

Thesis submitted to the Graduate Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

in

Dairy Science (Genetics)

APPROVED:

W. E. Vinson, Chairman

M. L. McGilliard

J. M. White

R. G. Cragle, Department Head

June, 1978

Blacksburg, Virginia

ACKNOWLEDGEMENTS

The author greatly appreciates the guidance, assistance and advice received from Dr. William Vinson in the planning and execution of this experimental work and in the preparation of this thesis.

The author also wishes to thank Dr. John White, Dr. Mike McGilliard, and Dr. R. G. Cragle for serving as members of the graduate committee.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	ii
LIST OF TABLES	v
LIST OF FIGURES.	vii
INTRODUCTION	1
REVIEW OF LITERATURE	3
Heterogeneity of Sire Group Variances for Production	
Traits	3
Possible Sources of Heterogeneity of Sire Variances.	4
Value of Repeatability	6
Cost of Repeatability.	8
DATA AND METHODS	10
Data Source.	10
Sire Daughter Variability.	12
Heterogeneity of Sire Daughter Variance.	14
Sire and Herd Effects.	17
Repeatability of Sire Daughter Variance.	17
Heritability of Daughter Variation	19
Within Sire Variability of Herd-Year-Season Daughter	
Averages	19
Relationship of Daughter Variability to Sire Repeatability .	21
PD Changes from Early to Later Proof	23
PD Changes of AI and and Non-AI Sires.	24
Value of R in Predicting Later Predicted Difference.	24
Cost of Repeatability.	25
RESULTS AND DISCUSSION	27
Sire Daughter Variability and its Heterogeneity Across Sires	27
Partitioning Sire Daughter Variability into Sire and Herd	
Effects.	35
Within Sire Variability of Herd-Year-Season Daughter	
Averages	42
Value of Repeatability	44
Cost of Repeatability.	56
CONCLUSIONS.	67
BIBLIOGRAPHY	69

Page

VITA 72

ABSTRACT

LIST OF TABLES

Table	Page
1. Means, standard deviations (SD) and ranges of sire variances.	29
2. Correlation coefficients between means and variances for three measures of milk yield	30
3. Sire and daughter factors contributing to the heterogeneity of sire daughter DAM variances	32
4. Contemporary factors contributing to the heterogeneity of sire daughter DAM variances.	34
5. Sire, daughter and contemporary factors contributing to heterogeneity of sire daughter DAM variances	36
6. Components of variance for sire daughter DAM variance 1964-1967.	37
7. Components of variance for sire daughter DAM variance 1967-1969.	38
8. Repeatability (t) of sire daughter DAM variance.	40
9. Heritability (h^2) of sire daughter DAM variance.	41
10. Within-sire standard deviations (SD) of herd-year-season daughter averages.	43
11. Within-sire standard deviations (SD) of herd-year-season daughter averages deviated from sire predicted difference.	45
12. Daughter variability by repeatability (R) class.	46
13. Means, standard deviations (SD) and ranges of PD changes with late repeatability ≥ 90	48
14. Regression of PD change on initial repeatability	49
15. Means, standard deviations (SD) and ranges of PD change with late repeatability (R) unrestricted	50
16. PD changes for AI and non-AI bulls	51
17. Prediction of final PDM (repeatability (R) ≥ 90)	54

Table	Page
18. Prediction of final PD\$ (repeatability (R) \geq 90).	55
19. Characteristics of 215 Hostein bulls with only production information	57
20. Regression coefficients and F values for semen cost model examining bulls with only production information.	58
21. Characteristics of 94 Holstein bulls with both production and type information.	59
22. Regression coefficients and F values for semen cost model examining bulls with both production and type information .	62

LIST OF FIGURES

Figure	Page
1. Frequency of standard deviations of daughter DAM	28
2. Final PD milk prediction from initial repeatability (RPT) and PD milk.	53
3. Semen cost prediction from repeatability (RPT) and PD milk with PD type constant.	61
4. Semen cost prediction from PD type and PD milk with repeatability (RPT) constant.	64
5. Semen cost per unit last PD milk (LPD) with PD type constant	65

INTRODUCTION

Within sire daughter variance of milk yield is an accepted, but often misunderstood, result of dairy sire selection and use. Bull production measures are usually presented in terms of "expected" production superiority or inferiority of the bull's future daughters in a breed average environment. The magnitude of possible deviations from such "expected" values are less well defined, but it is generally emphasized that large deviations can occur for individual daughters of any bull. User confidence in estimated bull production measures may improve if the expected magnitude of these deviations were known. Production from future daughters of bulls may then be expected to occur within a range of values instead of at a particular value. Heterogeneity of within sire variances has also been investigated. The specific purposes of the first part of this study were:

1. To determine the magnitude of expected deviations from predicted future daughter production.
2. To confirm heterogeneity of within sire daughter milk yield variability and investigate possible causes.

Repeatability is the accepted measure of the accuracy for individual estimates of dairy sire genetic merit. Repeatability is related to the rate of genetic improvement and is directly included in computing estimated sire merit. But there is little or no objective evidence that repeatability has additional value (over and above its use in computing estimated sire merit) to users of sire summary information. In spite of this, commonly quoted guidelines discourage dairymen from using bulls with less than 50-70% repeatability. Such guidelines tend

to minimize the use of high merit-low repeatability bulls and, in effect, advocate a system of sire selection which establishes independent culling levels for both merit and repeatability. For this system to yield optimum genetic or economic results, a value for repeatability must exist, which is supplemental to that accounted for in the theoretical relationship between true and estimated sire merit. The concept of a supplemental value for repeatability may have originated from experiences with the old national sire evaluation procedures. If so, any validity that the concept may have had would now be questionable since, with new sire evaluation procedures, both merit and repeatability are computed in entirely new ways.

The specific purposes of the second part of this study were:

1. To test several objective utility measures for increased accuracy (repeatability) of sire selection.
2. To determine a value for increased repeatability in terms of predicted difference.
3. To determine the cost of increased repeatability to dairymen.

REVIEW OF LITERATURE

Heterogeneity of Sire Group Variances for Production Traits

Several early researchers reported the existence of heterogeneity of variances within groups of paternal half sisters for production traits and examined the possibility of a positive relationship between the sire group variability and effects contributed by the sire. Johnson (1945) found significant differences among first lactation Holstein sire groups for variability of milk-fat yield in a daughter-dam comparison, but results of the study concluded that only 8% of the differences could be attributed to sires. Wadell, Van Vleck and Henderson (1960) reported heterogeneity of Holstein sire group variability for both milk and milk-fat yield for daughters of artificial insemination (AI) sires when the records were expressed as deviations from adjusted stablemate averages. Van Vleck and Barr (1963) reported highly significant differences among both first and second lactation Holstein sire groups for variability of milk yield records deviated from adjusted stablemate averages. Van Vleck (1968) reported significant heterogeneity of sire group variances of first lactation Holstein records expressed as deviations from adjusted stablemate averages. While the heterogeneity was not of sufficient magnitude to appreciably affect estimates of heritability, there was evidence that the observed differences could be attributed, at least partially, to effects of sires.

Van Vleck (1968) suggested that the repeatability of sire group variances from year to year should provide an estimate of the extent to which daughter variability is an effect of the sire. Repeatabilities

ranged from .39 to .63 for sire group variances expressed in six different ways. Sire effects thus accounted for 39% to 63% of the differences in sire group variances across years. Wadell, et al., (1960) reported intrasire regressions of second record variances on first record variances of .54 for milk and .52 for milk fat. Van Vleck and Barr (1963) reported product-moment correlations between first and second record deviations of .33 and .36 for sire group variances and natural logarithms of the variances.

Van Vleck and Barr (1963) reported that correlations between actual within group variance and estimated sire genotypic value (daughter-herdmate difference) were .20 for first lactations and .25 for second lactations. Correlations between variance and estimated genotypic value confounded with environmental level (daughter production) were .21 for first lactations and .22 for second lactations. These correlations indicated some relationship between the genotype for production and sire group variance, but the relationships accounted for little of the heterogeneity of variances.

Possible Sources of Heterogeneity of Sire Variances

Johnson (1945) suggested that the small differences in sire variances attributable to sire effects that he found were directly associated with the degree of inbreeding of the sire. However, Allaire and Henderson (1965) reported an estimated mean inbreeding level of .65% for 200 frequently used AI Holstein bulls. Only 4.4% of these bulls were inbred and the mean relationship among the sires and maternal grandsires of the inbred bulls was 16%. Henderson (1976)

reported that approximately 2% of the Holstein sires in Cornell University's Northeast AI sire comparison program were inbred and only one-eighth of these (.25% of total) were more than 12.5% inbred. Thus, the effect of inbreeding on sire group variance should be minimal.

Lee and Henderson (1964) reported heterogeneity of within-herd variances, which implies that some of the heterogeneity of sire group variance could be due to the particular herd in which the daughters make their records. Bereskin and Freeman (1965) reported that the components of variance of non-deviated milk yield records due to herds were about 32% of the total variance for both milk and milk-fat records. Components due to months, and the interactions of herd with months and year-seasons contributed 4.5 to 6.7% of the total variance of milk and milk-fat records. Burdick and McGilliard (1963) found little interaction between sires and herd environments classified by level of production, location, days dry, calving interval, or type of housing. Van Vleck (1968) stated that the simultaneous estimation of the effects of sire and herd on sire group variance would be impractical due to insufficient numbers of records in each sire by herd subclass.

Differences among genetic level of herdmates and genetic level of sires of herdmates may also be factors contributing to heterogeneity of sire group variances. McDaniel, Norman and Dickinson (1974) reported variation in genetic merit of sires of herdmates of first lactation cows by geographical region, AI organization and calendar year. Norman, McDaniel and Dickinson (1972) reported that as the average Predicted Difference (PD) for herdmates' sires in AI increased, so did

first lactation daughters' yield (regression coefficients = .35 to .73), but the daughters' deviation from AI herdmate average decreased (regression coefficients = -.72 to -1.14).

Differences in the level or variability of the genetic value of mates of sires may influence the variability of offspring production. Miller and Corley (1965) reported that, although there was a mean bias of 19 kg of milk per sire due to selection of mates, the relative ranking of Herdmate Comparison bulls was the same when daughter deviations alone were used as when both daughter and dam deviations were used for ranking. They found that the selection practiced in choosing cows to produce herd replacements did not result in substantial differences in variability between progeny and bull's mates. Intraherd standard deviations in milk yield were 941 kg and 930 kg for progeny and mates respectively. Keown and Kliever (1976) reported that, for PD values computed with the Modified Contemporary Comparison, the range in PD milk due to mates was 146 kg. Most sires were mated to dams that were above average, but the overall effect of dams was negligible.

Value of Repeatability

Repeatability is the accepted measure of accuracy for individual estimates of dairy sire genetic merit. Repeatability is related to the rate of genetic improvement and is directly included in computing estimated merit (PD). Dickinson, Powell and Norman (1976) reported that repeatability, as it is computed for the Modified Contemporary Comparison PD, accounts for number of daughters, number of herds, distribution of daughters across herds, number of records per daughter, length of

record, number of modified contemporaries per daughter, number of modified contemporaries' sires, and the average repeatability of modified contemporaries' sires. The repeatability is used in weighting both the modified contemporary deviated daughter milk yield and the bull's pedigree information.

There is little or no objective evidence that repeatability has additional value (over and above its use in computing estimated sire merit) to users of sire summary information. In spite of this, McKittrick (1976) recommended that dairymen use only bulls with repeatabilities greater than 70% and implied that low repeatability proofs are biased upwards. In response to a question concerning poor daughter performance from moderate PD sires, Mao (1976) suggested that low repeatability may be partially responsible for lower than expected daughter production. In another trade publication (Anonymous, 1976), dairymen were urged to place as much selection emphasis on repeatability as PD when selecting sires for AI use in their herds. The author stated that repeatabilities less than 70% indicated that the corresponding PD estimate was unreliable and that proofs with repeatabilities greater than 90% were very reliable. Wright (1975) recommended that dairymen avoid frequent use of any single low repeatability bull and that bulls with repeatabilities less than 50% be used very cautiously, if at all. McDaniel, Allaire and Dickinson (1968) stated that the best use of repeatability is as a guide to the extent to which a dairyman should use any individual bull.

Rothschild, Henderson and Everett (1976) reported that the Modified Contemporary Comparison appeared to remove a substantial portion of the biases favoring natural service sires in the Herdmate Comparison.

Cassell, et al. (1976) reported that bulls with a Herdmate Comparison limited-herd proof were overevaluated by 119 kg on their initial summaries compared with Herdmate Comparison multiherd proven bulls. A possible relationship between the limited-herd PD overevaluation and repeatability was not mentioned, but if a low repeatability identifies a limited-herd proof that is overevaluated, then repeatability should be considered in sire selection. Norman, Powell and Dickinson (1976) reported that the Modified Contemporary Comparison PD was more highly correlated (.22 versus .09) with additional daughter information than the Herdmate Comparison PD when fewer than 10 daughters were in the summary. The Modified Contemporary Comparison PD was only slightly more correlated (.73 versus .69) with additional daughter information than the Herdmate Comparison PD when more than 100 daughters were in the summary.

Cassell and Norman (1977) analyzed the accuracy of the Modified Contemporary Comparison to predict the performance of future daughters. Their analysis compared the accuracy of prediction of low and high PD levels and did not relate accuracy of prediction to repeatability of proof.

Cost of Repeatability

Gilmore, McDaniel and Carlson (1976) reported that dairymen avoided the risk of using young, low repeatability bulls because they were willing to forego a possible higher return in order to receive a less variable return from higher repeatability bulls. Dairymen were not adverse to risk when choosing proven, higher repeatability bulls and they tended to underestimate the value or accuracy of a young sires' pedigree

information. Bell and McDaniel (1976) reported that predicted differences for milk, dollars and type, together with the repeatabilities of predicted differences accounted for about 65 - 70% of the variation in semen prices. Repeatability alone was not a significant factor in accounting for semen cost variability, although the interaction between repeatability and PD dollars was a significant factor. Palmer and Mao (1977) reported that repeatability accounted for 7% of the variability of semen sales for bulls with both production and type information. Repeatability accounted for 12% of the variability of semen price for bulls with only production information.

DATA AND METHODS

Data Source

Data used in this study were 305 day or less, twice daily milked, mature equivalent, first lactation records of Holstein cows calving between January 1, 1957 and December 31, 1973. A first lactation record was defined as the record of a cow calving at less than 36 months of age with no previous record in the data set, whose actual record was based on at least 30 days of production. All records were obtained from and met requirements established by the Animal Improvement Programs Laboratory, USDA for use in sire and cow evaluations (Dairy Herd Improvement Letter, April, 1974).

Estimates of transmitting ability (PD) for milk and fat yield were computed for all Holstein sires having 10 or more first lactation daughters calving in any year, 1964 through 1973. Yearly sire summaries included all previous and current year daughters of any sire being summarized. Computing procedures were those used in the USDA Modified Contemporary Comparison (MCC) Sire Summary (Dickinson, et al., 1976) with the following exceptions:

1. Only first lactation records were used. Consequently, the comparisons among lactation records made in the same herd-year-season were contemporary comparisons rather than modified contemporary comparisons. Contemporaries of a cow were first lactation daughters of other sires in the same herd, who calved within two months of the cow's month of calving.

2. The genetic base year was 1965. The PD values of sires having repeatabilities (R) greater than or equal to 90% in the 1965 summary year (68 bulls) established the genetic base. PD values of all bulls from succeeding summary years were adjusted to this genetic base to allow direct comparison among all summaries. The adjustment was made as in the USDA MCC sire summary (Dickinson, et al., 1976).

3. Pedigree information in PD estimates used bull "group averages" determined by linear regression of daughter yield on pedigree index rather than actual computation of daughter yield (Dickinson, et al., 1976). Bull group averages defined the average genetic merit of bulls with similarly valued pedigrees. Weighting a bull's group average by $(1-R)$ incorporated available pedigree information in PD estimates and added credence to summaries of young bulls lacking an abundance of daughters' information.

4. Summaries were computed yearly, rather than three times per year.

MCC is an iterative procedure in which repeated estimates of genetic merit are used to adjust for genetic merit of sires of contemporaries. Estimates of genetic merit are updated and become more accurate in each succeeding iteration. The procedures required an initial estimate of contemporary sire genetic merit for use in the first summary year. The Herdmate Comparison sire summary procedures (Plowman and McDaniel, 1968) were used to compute repeatability and PD estimates of sires having 20 or more first lactation daughters calving between January 1, 1957 and December 31, 1973. These Herdmate Comparison PD values of 2597 Holstein bulls became the initial estimate of contemporary sire

genetic merit used in the 1964 Contemporary Comparison sire summary. Contemporary Comparison PD values were substituted for Herdmate Comparison PD values when the sire had sufficient 1964 through 1973 daughters to be summarized. A total of 992,409 records were used in calculating Contemporary Comparison PD values.

Sire Daughter Variability

The magnitude of within-sire daughter milk yield variance has important implications in the interpretation, use and acceptance of sire evaluation procedures by dairymen. Sire PD values are usually interpreted in terms of "expected" production superiority or inferiority of the sire's future daughters in breed average herds. The probability of sizable deviations from this expected production is emphasized but the magnitude of the deviations is not well defined. Unexpected deviations of daughter production from the level predicted for her sire may erode user confidence in sire evaluation procedures. Within-sire daughter variances were therefore computed to quantitate the probabilities of given sized deviations from expected production for single daughters and groups of daughters. These values were compared to expected within-sire daughter variances computed from the following formula:

$$\sigma_{D_{ij}}^2 = \left[\frac{n_{ij} + 14}{16n_{ij}} + \frac{15}{16m_{ij}} \right] \sigma_w^2 \quad [1]$$

where,

$$\sigma_{D_{ij}}^2 = \text{variance of the } j\text{th sire's mean daughter milk yield in herd-year-season } i, \text{ deviated from mean contemporary milk yield and adjusted for contemporary sire genetic merit,}$$

n_{ij} = number of daughters of sire j in herd-year-season i ,

m_{ij} = number of contemporaries of the daughters of sire j in herd-year-season i ,

and σ_w^2 = within herd-year-season variance of daughter milk yield deviated from mean contemporary milk yield and adjusted for contemporary sire genetic merit (DAM).

This formula is derived assuming a within herd-year-season heritability for milk yield of .25, such that the within sire herd-year-season variance of yield is $15/16 \sigma_w^2$. It is further assumed that there is no residual environmental covariance among daughters or contemporaries within herd-year-seasons, and that each contemporary is by a different sire. The variance of mean daughter yield therefore includes a term which is inversely related to the increase in daughter number (the reduction is not proportional to the increase in daughter number due to the genetic relationship among daughters) and a term which is inversely proportional to the number of contemporaries in the contemporary average. Cows calving between February 1964 and November 1967 and included in the 1967 sire summary year (209,315 cows) were used to estimate σ_w^2 . Herd effects were absorbed and DAM was then regressed on year of calving. The error mean square from the regression procedure estimated within herd-year variance. Seasonal variation in deviated records was assumed sufficiently small to be ignored (Bereskin and Freeman, 1968). An identical procedure using all cows from the 1972 sire summary year (456,708 cows) allowed a comparison of within herd-year variances to determine the change in variance with time.

Variances of daughter milk yield (M), daughter milk yield deviated from mean contemporary milk yield (DM), and DAM were computed for all bulls having at least 10 daughters and at least 10 mates with production records available for the 1967 sire summary year. The number of contemporary sires with Herdmate Comparison PD values in the 1967 sire summary year was assumed to be sufficiently small that the Contemporary Comparison PD computations were not appreciably affected. Variances were computed for 1574 bulls with a total of 205,361 daughters. Means, standard deviations and ranges of sire variances for the three measures of milk yield were calculated and sire variances were tested for heterogeneity using Bartlett's test (Bartlett, 1937). The distribution of standard deviations of daughter DAM was graphed. Correlations between means and variances for the three measures were computed and tested for significance.

Heterogeneity of Sire Daughter Variance

The detection of significant ($P < .01$) levels of heterogeneity of within-sire daughter variance resulted in several attempts to identify general and specific factors associated with the sizable differences in sire daughter variability observed. For each of the 1574 bulls, the following information was computed: mean daughter milk yield, mean and variance of contemporary milk yield, mean and variance of contemporary lactation length, mean and variance of contemporary sire R, mean and variance of contemporary sires' daughter DM, mean and variance of number of contemporary sires, number of daughters, number of different daughter herds, mean daughter DAM, number of different months in which daughters

calved, mean and variance of estimated relative producing ability (ERPA) of mates, PD milk and R. Mates' ERPA values were based upon herdmate deviated milk yield since contemporary information was unavailable.

ERPA values were calculated using the following formula:

$$\text{ERPA} = W(\text{Average Herdmate Deviation})$$

where,

$$W = \frac{.5 (\text{number of records})}{1 + .5 (\text{number of records} - 1)}$$

Each sire's variance of daughter DAM (VDAM) was converted to common logarithms and analyzed by regression procedures to identify factors associated with differences in sire variance. A total of 32 models containing various combinations of the 19 factors listed above were tested. The models tested were compared and three "best" models were selected that most effectively accounted for differences in sire daughter variances and included similar factors (contemporary factors versus sire and daughter factors).

Model 1 included factors unique to the sire and his daughters.

Model 1 was

$$\begin{aligned} \text{Log(VDAM)} = & b_1 \bar{M} + b_2 \bar{M}^2 + b_3 \bar{M}^3 + b_4 R + b_5 \overline{\text{DAM}} + \\ & b_6 \text{VERPA} + b_7 \text{MON} + b_8 \text{MON}^2 \end{aligned}$$

where,

$$\bar{M} = \text{mean daughter milk yield,}$$

R = sire's repeatability,
 \overline{DAM} = mean daughter DAM,
 $VERPA$ = variance of mates' ERPA,
 and, MON = number of different months in which the
 sire's daughters calved.

Model 2 consisted of factors unique to contemporaries of the sire's daughters. Model 2 was

$$\text{Log}(VDAM) = b_1 \overline{CM} + b_2 VCM + b_3 NCS + b_4 \overline{CSDAM} + b_5 VCSR$$

where,

\overline{CM} = mean contemporary milk yield,
 VCM = variance of contemporary milk yield,
 NCS = number of contemporary sires,
 \overline{CSDAM} = contemporary sires' mean daughter DAM,
 and, $VCSR$ = variance of contemporary sires' repeatability.

Model 3 was a combination of Models 1 and 2 and included daughter, sire and contemporary factors. Model 3 was

$$\text{Log}(VDAM) = b_1 \overline{M} + b_2 \overline{M}^2 + b_3 \overline{M}^3 + b_4 R + b_5 \overline{DAM} + b_6 VERPA + b_7 MON + b_8 MON^2 + b_9 \overline{CM} + b_{10} VCM + b_{11} NCS + b_{12} \overline{CSDAM} + b_{13} VCSR$$

where, all factors are as previously defined. This model allowed a comparison of the relative importance of the daughter, sire and

contemporary factors.

Sire and Herd Effects

The general effects of sires and herds on within-sire daughter variance were examined using Henderson's (1953) Method I analysis of variance on daughters included in the 1967 sire summary year (calving dates were March 1964 through October 1967). Sire by herd subclass variances of DAM were calculated based on groups of ten random daughters of the same sire in the same herd. More than one variance for a particular sire and herd was possible. Sires with only one sire by herd subclass variance were eliminated from the subsequent analysis to maximize the number of filled sire by herd subclasses. An identical examination of the effects of sires and herds on daughter variance for a data set consisting of cows calving from November 1967 through October 1969 was performed in an attempt to increase the number of filled subclasses and afforded a test for consistency of sire and herd effects.

Repeatability of Sire Daughter Variance

The purpose of the sire and herd analysis was to estimate the percentage of differences in sire daughter variation which may be attributed to the sire. However, it was anticipated that the high frequency of missing sire-by-herd subclasses likely to occur would make conclusions drawn from this analysis subject to question. It was therefore determined to compute the repeatability of sire daughter variance across years as a further estimate of the percent of differences in daughter variability which are characteristic of the particular sire.

The repeatability of sire variance of daughter DAM across years was estimated from a within and among sire nested analysis of variance. Repeatability of sire variance was estimated using only those sires that had at least 25 daughters in any one year and, then using only those sires that had at least 50 daughters in any one year. Cows calving between October of the preceding year and November of the given year were included in the variances for each year. Sire variances of DAM were computed for the years 1967 through 1973. A total of 3074 sires had at least 1 variance of 25 daughters and a total of 1930 sires had at least 1 variance of 50 daughters.

In an attempt to determine the effect of herds on the repeatability of sire variability, a second data set was formed in which a completely different set of herds was required for a given sire in each of the different years. The repeatability of variance of daughter DAM may be affected by the type of herds in which a sire is used or by a confounding effect between sires and herds. Deciding upon an acceptable set of criteria upon which to base decisions on "type of herd" proved difficult, consequently only those cows which were in completely different herds from their paternal half-sibs of previous years were included in later variances. Variances of daughter DAM of those sires that had at least 25 daughters in any one year for the years 1967 through 1973 were used in a within and among sire nested analysis of variance to determine the repeatability of sire variability. A total of 1581 sires had at least 1 variance of 25 daughters.

Heritability of Daughter Variation

A non-zero heritability of daughter variation allows the estimation of the variability of future daughters of a young sire from ancestral information. Variances of DAM calculated from groups of 25 or more daughters and from groups of 50 or more daughters were used in a regression of son on sire to estimate heritability. Doubling regression coefficients yielded heritability estimates from the first available (initial) yearly variance of both sire and son and from a total variance of both sire and son. A regression of son's initial variance on sire's total variance produced an estimate of daughter variation heritability which could be more practically useful. A total of 500 sire-son pairs (127 different sires) had a minimum of 25 daughters and 258 sire-son pairs (75 different sires) had a minimum of 50 daughters.

Within Sire Variability of Herd-Year-Season Daughter Averages

To determine the amount of variability in mean milk yield that may be expected from groups of daughters of a single sire in breed average herds, groups of 2, 4, 6, 8, 10, 12, 14 and 16 daughters of the same sire in the same herd-year-season (H-Y-S) were obtained from cows used in the 1972 sire summary year. Only one group of daughters of a particular group size was allowed in a H-Y-S, but smaller sized groups were obtained from the larger sized groups. For example, if there were 12 daughters of a sire in a particular H-Y-S, groups of 10, 8, 6, 4, and 2 daughters were also obtained from that group of 12 daughters.

The mean DAM of the i th group of the j th sire (\bar{D}_{ij}) was obtained

for each group and deviated from the mean DAM for all equal-sized groups of daughters of that same sire (G_j). All $(\bar{D}_{ij} - G_j)$ were grouped according to number of daughters in the mean and regressed on the reciprocal of the number of contemporaries of the group to remove the effects of differing numbers of contemporaries across sires (see formula 1). The standard deviations (square root of error mean square) of $(\bar{D}_{ij} - G_j)$ were compared with predicted standard deviations determined from formula 1 which assumes contemporaries to be unrelated. Allowing for covariances (genetic or environmental) among contemporaries would increase predicted standard deviations.

The \bar{D}_{ij} for each group of 1968 through 1972 daughters was also deviated from their sire's 1967 PD ($\bar{D}_{ij} - PD_j$), such that PD values and daughter averages were determined from independent samples of daughters. The $(\bar{D}_{ij} - PD_j)$ values were categorized according to group size and regressed on the reciprocal of number of contemporaries of each group. Standard deviations (square roots of error mean square) of $(\bar{D}_{ij} - PD_j)$ adjusted for differences in contemporary numbers were compared with predicted standard deviations determined from formula 1 which was modified to account for less-than-perfect sire's R. Assuming within H-Y-S heritability of .25, the modified equation was

$$\sigma_{\bar{D}_{ij}}^2 = (1-R)\sigma_s^2 + \left[\frac{n_{ij} + 14}{16n_{ij}} + \frac{15}{16m_{ij}} \right] \sigma_w^2 ,$$

$$\sigma_{\bar{D}_{ij}}^2 = \left[(1-R) \cdot 0625 + \frac{n_{ij} + 14}{16n_{ij}} + \frac{15}{16m_{ij}} \right] \sigma_w^2$$

where σ_s^2 is variance due to differences among breeding values of sires ($\sigma_s^2 = .0625 \sigma_w^2$) and all other terms are as previously defined. Actual and predicted standard deviations of $(\bar{D}_{ij} - PD_j)$ were expected to be larger than the standard deviations of $(\bar{D}_{ij} - G_j)$ for equal-sized groups of daughters. Since the sire's value (G_j) is held constant for variance of $(\bar{D}_{ij} - G_j)$, there is no variance due to differences in true sire additive genetic values. Therefore, sires do not contribute to the variance of $(\bar{D}_{ij} - G_j)$ since G_j is constant. However, the variance of $(\bar{D}_{ij} - PD_j)$ is affected to the extent that PD_j does not perfectly estimate true transmitting ability.

For each analysis, predicted values were determined assuming homogeneity of error variance (within-sire herd-year-season variance). Heterogeneity of sire daughter variance implies that the variance of daughter records differs across sires. For bulls having less daughter variance than assumed, the expected within-sire daughter standard deviations should be smaller than the predicted values. For bulls having more daughter variance than assumed, the expected standard deviations should be larger than the predicted values.

Relationship of Daughter Variability to Sire Repeatability

As a group, daughters of high R sires are theoretically expected to be less variable than the daughters of low R sires, when estimated sire merit is held constant. The reduced variability arises from a reduction or removal of sire true genetic differences from the observed daughter variation, and can amount to 1/4 of the additive genetic variance when heritability is .25. That is, daughters of bulls whose

breeding values are known ($R=1.00$) may be expected to be 1/16 or 6.25% less variable than the daughters of bulls about whom nothing is known ($R=0$).

The result for cases intermediate to these extremes is obtained assuming true sire transmitting abilities and their estimates are bivariate normally distributed with variance $\sigma_s^2 (1-R)$. Then

$$\sigma_{w^*}^2 = (1-R)\sigma_s^2 + 3/4 \sigma_A^2 + \sigma_E^2$$

where,

$\sigma_{w^*}^2$ = total within H-Y-S phenotypic variance of daughter yield, holding estimated sire genetic merit constant,

R = sire's repeatability,

σ_s^2 = variance due to differences among breeding values of sires,

σ_A^2 = additive genetic variance,

and, σ_E^2 = variance due to all remaining causes.

The sire's portion of total variance can be converted to $1/4 (1-R)\sigma_A^2$

and the equation becomes

$$\sigma_{w^*}^2 = 1/16 (1-R)\sigma_w^2 + 15/16 \sigma_w^2 \quad [2]$$

where σ_w^2 is total within H-Y-S phenotypic variance and all other terms are as previously defined.

Cows calving between October 1966 and November 1967 whose sires had at least 10 daughters (73,960 cows) were used to examine the relationship between daughter variance and sire R . To adjust for sire's PD,

daughter's DAM was regressed on sire's PD. The residuals of DAM from the regression procedure were then grouped by sire's R and the variances of the residuals were examined by R group. The decrease in daughter variance per 1% increase in sire R was calculated and the percent decrease in daughter variance over the range of repeatabilities was compared with an expected value determined from equation 2.

PD Changes from Early to Later Proof

One measure of a supplemental value for R was the reduced probability of random overestimations of genetic merit. If the PD values of low R proofs have a greater probability of declining than the PD values of high R proofs, then selection for high R is justified. Using the Contemporary Comparison PD values from the yearly sire summaries, the first available (initial) post-1966 PD was deviated (late PD minus early PD) from the latest available PD that had a corresponding R of 90% or greater (455 bulls). It was assumed that the 90% R proof accurately estimated the sire's genetic merit and that earlier deviations from that estimate were random errors for AI sires. The changes in PD values from initial to late proof were grouped according to sire's initial R and examined by R group. The change in PD was then regressed on sire's initial R. The negative and positive PD changes were then separately regressed on bull's initial R. These procedures were also performed on a data set containing proofs of bulls whose latest R was not required to be equal to or greater than 90% (2086 bulls). The latter data set increased the number of proofs in the R classes and was compared with the previous data set.

PD Changes of AI and Non-AI Sires

Another measure of a supplemental value for R was the reduced probability of non-random overestimation of sire merit. Repeatability may have a value as an indicator of possibly biased non-AI proofs. Bulls with non-AI proofs were identified as bulls with daughters in four or less herds whose earliest AI daughter¹ calved in some later year. Any proof not defined to be a non-AI proof was assumed to be an AI proof. The first available (initial) post-1966 non-AI proof was deviated from the latest available proof of 90% or greater R. It was assumed that the 90% R proof accurately estimated a sire's genetic merit and that earlier deviations from that estimate were a combination of random and non-random errors. Deviations for AI proofs were random errors and the differences between average deviations for AI and non-AI proofs were considered to be non-random errors. The results were examined by sire's early R group.

Value of R in Predicting Later Predicted Difference

A specific supplemental value for R in terms of predicted difference was desired to allow comparison of proofs of bulls with differing predicted differences and R. Adjusting sires' predicted differences for R would eliminate the need for arbitrary guidelines for acceptable R which tends to minimize the use of low R sires. All bulls having both an initial post-1966 proof and a later proof with an R of 90% or

¹As determined from NAAB Cross Reference File obtained from Animal Improvement Programs Laboratory, USDA, Beltsville, MD.

greater (455 bulls) were used in regression procedures to quantitate this supplemental value for R. A proof with R of 90% was assumed to be an accurate estimate of the bull's true genetic value. Each bull's later predicted difference milk (PDM) was regressed on his initial PDM and initial R, separately and in combination, to estimate the absolute and relative values of these initial measures of the sire in predicting his final, or true, transmitting ability. The possibility of non-linear relationships was examined by including in the model quadratic effects of initial PDM and R. Identical analyses were performed on predicted difference dollars (PD\$). Predicted difference dollars was calculated using the following formula:

$$\text{PD\$} = \text{PDM}(.2095 - 2.5132(.035)) + \text{PDF}(2.5132)$$

where,

.2095 = price paid for a kg of milk at the base test,

2.5132 = price paid for a kg of fat,

.035 = base test,

PDF = predicted difference fat.

Cost of Repeatability

The optimum emphasis afforded R in sire selection should be a function of the cost of R as well as its supplemental value. It is logical to measure the former in terms of the relationship between R and semen price. To determine the effect of R on semen cost, commercial semen price lists for all available Holstein bulls were requested from ten major bull stud centers. The nine bull stud centers which supplied

semen cost information accounted for 92% of 1974 U. S. dairy semen sales, and thus represent most of the industry. All semen prices were as of July 1, 1977 for a single service. Spring 1977 production proofs were obtained from the Animal Improvement Programs Laboratory, USDA, Beltsville, Maryland. Type proofs for 1977 were obtained from the Holstein-Friesian Association of America, Brattleboro, Vermont. Several regression models were fit to identify major factors affecting semen prices. Factors examined were stud, PDM linear and quadratic, R linear and quadratic for PDM, predicted difference fat percent (PDFF), PD\$ linear and quadratic, predicted difference type (PDT) linear and quadratic, R for PDT, and interaction between linear terms. Only bulls with PDM exceeding 100 kg and semen cost less than \$40 per unit were examined. A total of 94 bulls with both production and type evaluations were available for analysis. Separate analyses were performed on 215 bulls with production evaluations only.

RESULTS AND DISCUSSION

Sire Daughter Variability and its Heterogeneity Across Sires

Within herd-year variance (σ_w^2) estimated from DAM values of cows calving from 1964 through 1967 was 1,814,171 kg² and from DAM values of cows calving from 1964 through 1972 was 1,937,062 kg². These values were larger than within herd-year-season variances estimated by earlier researchers. Within herd-year variance of DAM increased approximately 6.8% over five years for contemporary comparison records.

Means, standard deviations and ranges of sire daughter variances are presented in Table 1 and differed only slightly for the three measures of daughter milk yield. However, very large and highly significant ($P < .01$) differences were found among sires for daughter variability. For example, daughter variability for individual sires ranged from 187,291 to 5,261,858 kg² for DAM. The distribution of daughter DAM standard deviations for 1574 bulls is presented in Figure 1. The figure illustrates the frequency of bulls with daughter standard deviations of DAM within a specific number of kgs of the mean standard deviation for all bulls. The mean standard deviation was 1260 kg and the standard deviation of the individual estimates was 232 kg. Correlations between means and variances for the measures are presented in Table 2. This highly significant correlation between the mean and variance of unadjusted daughter milk yield (D) suggests the presence of a scaling effect which is removed when daughter milk yield is adjusted for mean contemporary milk yield and mean contemporary sire genetic merit (DM and DAM).

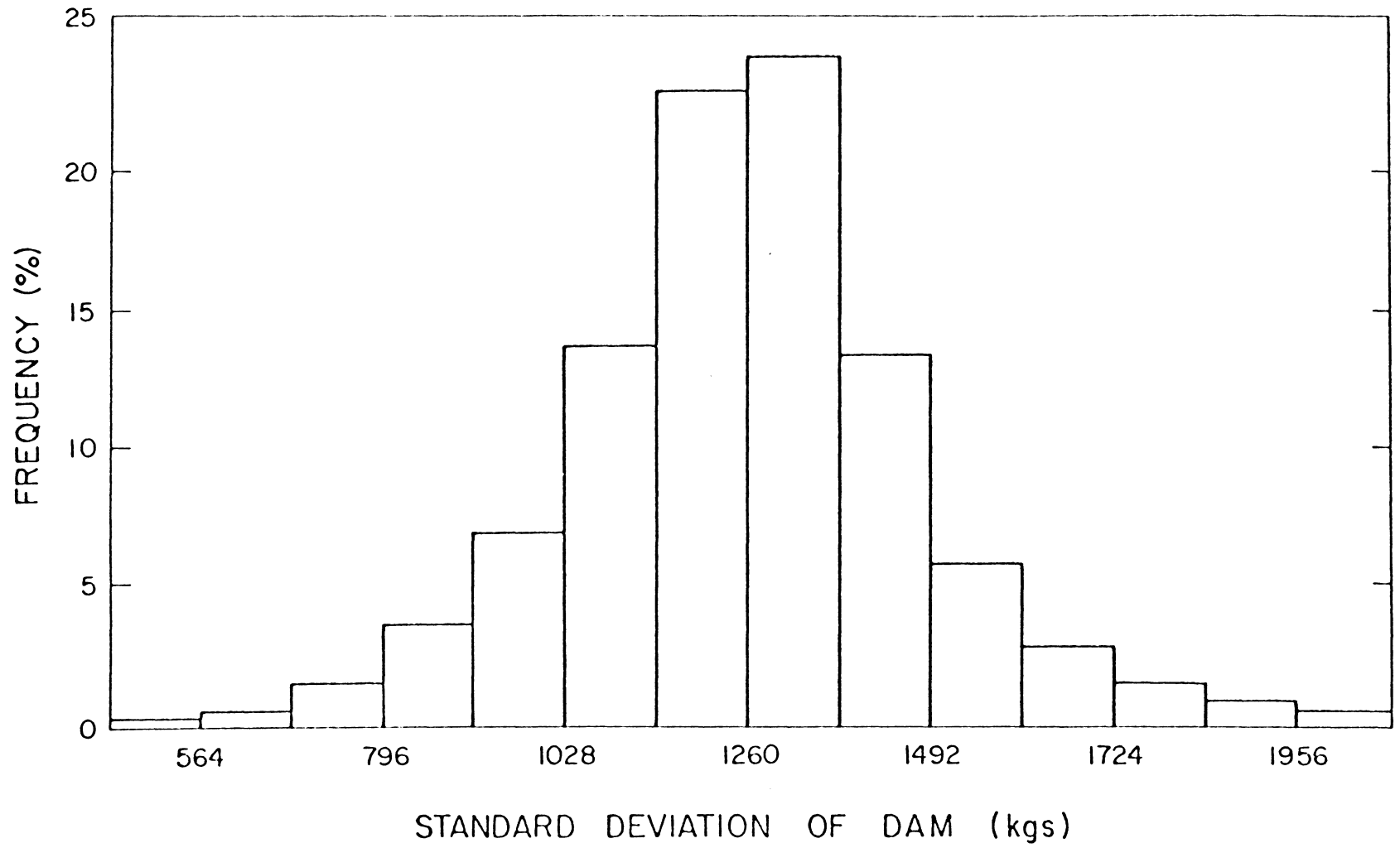


FIGURE 1. FREQUENCY OF STANDARD DEVIATIONS OF DAUGHTER DAM.

TABLE 1. MEANS, STANDARD DEVIATIONS (SD) AND RANGES OF SIRE VARIANCES.

Measure	No. of Sires	Mean	SD	Range	Chi-Square
M	1574	1,719,058	610,466	146,012 to 5,409,717	3812.5**
DM	1574	1,670,516	626,325	194,061 to 5,535,272	3470.3**
DAM	1574	1,642,006	611,938	187,291 to 5,261,858	3434.1**

** Significant ($P < .01$)

TABLE 2. CORRELATION COEFFICIENTS BETWEEN MEANS AND VARIANCES FOR
THREE MEASURES OF MILK YIELD

<u>MEASURE</u>	<u>CORRELATION</u>
M	.252**
DM	-.012
DAM	-.018

** Significant (P<.01)

Standard partial regression coefficients and F values for sire and daughter factors contributing to observed differences in logarithms of sire variances for DAM are presented in Table 3. The significant linear, quadratic and cubic effects of unadjusted daughter milk yield suggests the presence of a non-linear scaling effect. However, this was not a scaling effect in the usual sense, due to an absence of a correlation between mean and variance, and the fact that the effect was measured on log transformed data. In spite of the smaller F values, the standard partial regression coefficients imply a greater relative importance for these three factors in determining sire variance than the other factors considered. The number of different months in which the sire's daughters calved (MONTHS) and MONTHS quadratic significantly affected sire daughter variance. MONTHS measures the effect of time span of daughter records on daughter variability. An increase in sire's R was associated with a significant increase in daughter variability. This was probably due to the positive relationship between R and the number and distribution of daughters. When sire's R and MONTHS were considered together, higher R and wider span of calving months were apparently associated with greater daughter variability. However, when actual regression coefficients are considered, the significant negative quadratic effect of MONTHS causes a reversal of this effect when MONTHS is greater than 51. The highly significant effect of mean DAM of the sire's daughters implies that adjustment for mean contemporary milk yield and contemporary sire genetic merit accounted for daughter variability in addition to that explained by unadjusted daughter milk yield. Although the variance of mates' ERPA values had the greatest F value in

TABLE 3. SIRE AND DAUGHTER FACTORS CONTRIBUTING TO THE HETEROGENEITY
OF SIRE DAUGHTER DAM VARIANCES¹.

FACTOR	STANDARD PARTIAL REGRESSION COEFFICIENT	F VALUE
\bar{M}^2	-7.25849	5.11*
\bar{M}	4.06356	5.94*
\bar{M}^3	3.45502	4.86*
Months	0.29424	5.42*
Months ²	-0.28019	6.20*
Sire R	0.15302	8.65**
Sire $\overline{\text{DAM}}$	-0.15207	29.56**
V (Mates' ERPA)	0.13745	30.81**

¹. Multiple squared correlation = .11.

*Significant (P<.05)

**Significant (P<.01)

this model, its relative importance in accounting for daughter variability (standard partial regression coefficient) was smallest. This may be partially due to the fact that little variability in the variance of mates' ERPA values, compared to other factors considered, was observed in the data examined. This model accounted for 11% of the differences in sire variances.

The second model contained factors unique to daughters' contemporaries and is presented in Table 4. The variance of contemporary milk yield was the most significant factor and its standard partial regression coefficient indicated that as the variance of contemporary milk yield increased, so did daughter variability. The mean contemporary milk yield also had a positive significant effect on daughter variability. The latter is probably a scaling effect, but both results are expected where daughter performance is expressed as a deviation from average contemporary yield. The number of contemporary sires and the variance of contemporary sire R were significantly and positively associated with sire daughter variance. An increase in the former may be an indirect indicator of increasing herd size in which greater daughter variation might be expected. The latter relationship is expected from the use of contemporary deviated records since repeatability had a small but significant effect on daughter variability. The contemporary sire mean DAM was significantly and negatively associated with sire daughter variance. This may also reflect a scaling effect since, other things being equal, a lower contemporary average results in a higher daughter DAM record. This model accounted for 16% of the differences in sire daughter variances.

TABLE 4. CONTEMPORARY FACTORS CONTRIBUTING TO THE HETEROGENEITY OF
SIRE DAUGHTER DAM VARIANCES¹.

FACTOR	STANDARD PARTIAL REGRESSION COEFFICIENT	F VALUE
V (CMILK)	0.29565	144.82**
$\overline{\text{CMILK}}$	0.21025	78.09**
Number CS	0.12358	25.37**
V (CS R)	0.08635	11.85**
CS $\overline{\text{DAM}}$	-0.05517	5.55*

¹. Multiple squared correlation = .16.

*Significant (P<.05)

**Significant (P<.01)

The third model was a combination of the sire, daughter model and the contemporary model and is presented in Table 5. The combined model illustrated the relative importance of the sire, daughter, and contemporary factors. The linear, quadratic and cubic effects of unadjusted daughter milk yield was again associated with the greatest change in daughter variability and suggests the presence of a non-linear scaling effect not removed by logarithmic transformation. The variance of contemporary milk yield was also highly significantly and positively associated with sire daughter variance. The ranking of the remaining factors, by standard partial regression coefficients, changed little relative to other factors from the same initial model. This model accounted for only 18% of the differences in sire daughter variances. Thus, 82% of the differences in sire daughter variances could not be accounted for by the specific factors examined.

Partitioning Sire Daughter Variability into Sire and Herd Effects

Sire and herd components of variance for sire variance of cows calving from 1964 through 1967 are presented in Table 6. These indicate approximately equal contributions from sires and herds, and suggest that a considerable part of the heterogeneity of sire variances is associated with sires. However, since only .42% of the possible sire by herd subclasses were filled, this conclusion is made with caution. Similar conclusions can be made from the 1967 through 1969 data obtained in an effort to fill more subclasses (Table 7). The percent of total variance for sire and herd did not differ appreciably from those of the previous data set and were approximately equal. However, only .45% of the

TABLE 5. SIRE, DAUGHTER AND CONTEMPORARY FACTORS CONTRIBUTING TO
HETEROGENEITY OF SIRE DAUGHTER DAM VARIANCES¹.

FACTOR	STANDARD PARTIAL REGRESSION COEFFICIENT	F VALUE
\bar{M}^2	-6.63739	4.60*
\bar{M}	3.54052	4.86*
\bar{M}^3	3.10353	4.22*
V(CMILK)	0.27572	102.29**
Months	0.24844	4.14*
$\overline{\text{CMILK}}$	0.20501	4.14*
Months ²	-0.18216	2.80
V (Mates' ERPA)	0.11705	23.71**
Number CS	0.10723	18.68**
V (CS R)	0.07418	8.62**
CS $\overline{\text{DAM}}$	-0.05129	2.17
Sire R	-0.03252	.38
Sire $\overline{\text{DAM}}$	0.02771	.11

¹. Multiple squared correlation = .18.

*Significant (P<.05)

**Significant (P<.01)

TABLE 6. COMPONENTS OF VARIANCE FOR SIRE DAUGHTER DAM VARIANCE¹.
1964-1967

<u>SOURCE</u>	<u>df</u>	<u>COMPONENT</u>	<u>PERCENT OF TOTAL</u>
Herds	951	.0592	43.5
Sires	432	.0553	40.7
Error	349	.0215	15.8

¹.Maximum percent filled subclasses = .42.

TABLE 7. COMPONENTS OF VARIANCE FOR SIRE DAUGHTER DAM VARIANCE¹.
1967-1969

<u>SOURCE</u>	<u>df</u>	<u>COMPONENT</u>	<u>PERCENT OF TOTAL</u>
Herds	995	.05075	39.6
Sires	421	.05353	41.7
Error	484	.02399	18.7

¹.Maximum percent filled subclasses = .45.

possible subclasses were filled and these conclusions are also subject to question.

The repeatability of sire variance of daughter DAM across years was calculated to obtain a more accurate estimate of the percent of differences in daughter variability which may be attributed to sires. Repeatabilities are presented in Table 8 for three different data sets. Repeatability estimates for minimums of 25 and 50 daughters per sire, per year, with no restriction for different herds, were approximately equal, but were less than Van Vleck's (1968) estimate of .38 computed from a minimum of 50 daughters per sire, per year. These estimates were made with no attempt to specifically remove the effect of herds. Requiring daughters in different years to be in different herds resulted in an estimate of 23% for the proportion of differences in sire daughter variances due to sires. The differences between this estimate and the two previous estimates may be attributed to a partial confounding of sire and herd effects across years. The repeatability of sire variances yielded lower estimates of the portion of daughter variability attributable to the sire than did estimates from the sire by herd analysis. The repeatability of sire daughter variance estimates was considered to be a more accurate method of determining the sire's proportion of daughter variability due to the extreme unbalanced nature of the sire by herd analysis.

Heritability estimates of daughter variation are presented in Table 9. With a minimum of 25 daughters per variance, heritability calculated from first available (initial) estimates of daughter variance for both sire and son were negative, suggesting a true value near zero.

TABLE 8. REPEATABILITY (t) OF SIRE DAUGHTER DAM VARIANCE.

<u>NO. OF DAUGHTERS</u> <u>PER YEARLY VARIANCE</u>	<u>NO. OF</u> <u>SIRES</u>	<u>NO. OF</u> <u>VARIANCES</u>	<u>t</u>
25	1127	3074	.29
25 ^{1.}	719	1581	.23
50	693	1930	.32

^{1.} Daughters in different years are in different herds.

TABLE 9. HERITABILITY (h^2) OF SIRE DAUGHTER DAM VARIANCE.

NO. OF DAUGHTERS PER YEARLY VARIANCE	VARIANCE TYPE ¹ .	NO. OF SIRE-SON PAIRS	NO. OF SIREs	h^2
25	initial-initial	500	127	-.14 _± .13
25	total-total	500	127	.22 _± .12
25	total-initial	500	127	.08 _± .14
50	initial-initial	259	75	.11 _± .13
50	total-total	259	75	.29 _± .11
50	total-initial	259	75	.21 _± .13

¹. Variance listed first is sire's, that listed second is son's.

The remaining heritability estimates were low to moderate, with relatively large standard errors. Use of the more accurate estimates for sire and son daughter variability (total variance) resulted in higher heritability estimates. Heritability decreased when son's initial daughter variance was regressed on sire's total daughter variance. The heritability estimates indicate that if the daughters of a sire have highly variable DAM values, then the daughters of the sons of that sire are more likely to have highly variable DAM values.

Within Sire Variability of Herd-Year-Season Daughter Averages

Actual and predicted within-sire standard deviations of H-Y-S daughter averages for $(\bar{D}_{ij} - G_j)$ are presented in Table 10. Tabled values may be interpreted in terms of expected deviation from mean yield of equal-sized sire daughter groups. For example, if the group average DAM for an infinite number of normally distributed groups of 2 daughters were deviated from the mean DAM for all groups of 2 daughters of the same sire, 67% of the mean DAM values would be expected to be within 999 kg of the mean DAM for all groups of 2 daughters of that sire. Another 33% of the mean DAM values would be expected to be more than 999 kg above, or more than 999 kg below the mean DAM value for all groups. Actual standard deviations decreased as the number of daughters in the average increased in relatively good agreement with predicted values. Although a χ^2 test showed that there were highly significant ($P < .01$) differences between predicted and actual standard deviations, the predicted standard deviations were relatively accurate, with a maximum prediction error of about 10%.

TABLE 10. WITHIN-SIRE STANDARD DEVIATIONS (SD) OF HERD-YEAR-SEASON DAUGHTER AVERAGES.

DAUGHTERS PER MEAN	NO. OF MEANS	NO. OF SIREs	NO. CONTEMPOR- ARIES PER MEAN	SD	
				ACTUAL	PREDICTED ^{1.}
2	73172	1929	7.17	999	1105
4	11792	1632	9.85	816	854
6	3796	1230	11.95	745	745
8	1564	776	14.68	678	676
10	736	461	17.93	642	626
12	346	248	19.28	577	597
14	196	155	21.94	535	497
16	113	96	25.65	522	546

$$1.\text{Predicted SD} = \left[\sigma_w^2 \left[\frac{n_{ij} + 14}{16n_{ij}} + \frac{15}{16m_{ij}} \right] \right]^{1/2}$$

Table 11 shows the within-sire standard deviations of H-Y-S daughter group averages deviated from their sire's PD determined from previous daughters ($\bar{D}_{ij} - PD_j$). Actual standard deviations of ($\bar{D}_{ij} - PD_j$) were larger than the standard deviations of ($\bar{D}_{ij} - G_j$). This was expected since the former includes any errors in estimating PD_j . Standard deviations of ($\bar{D}_{ij} - PD_j$) decreased approximately as expected with increasing numbers of daughters in the \bar{D}_{ij} . Although a χ^2 test showed that there were highly significant ($P < .01$) differences between predicted and actual deviations, the maximum prediction error was 11%. The standard deviations indicate that milk yield of groups of daughters of the same sire calving in the same H-Y-S can vary widely from the predicted value (sire's PD) even when several daughters are included in the average. For example, when there are groups of 10 daughters in a H-Y-S, the mean DAM of only 67% of the groups of daughters are expected to be within 669 kg of the predicted value for the daughters of that sire.

Value of Repeatability

Daughter variability (after accounting for sire's PD) by R class is presented in Table 12. Variances decreased approximately 537 kg^2 per 1% increase in sire's R. From lowest to highest R class, this is a decrease of 2.25% compared to an expected decrease in daughter variance of 4.5% over this range of sire R. This small decrease in daughter variation did not indicate a supplemental value for R.

Another measure of a supplemental value for R was the reduced probability of random overestimations of genetic merit. Means, standard deviations and ranges of changes in PD values from early to final proof

TABLE 11. WITHIN-SIRE STANDARD DEVIATIONS (SD) OF HERD-YEAR-SEASON DAUGHTER AVERAGES DEVIATED FROM SIRE PREDICTED DIFFERENCE.

DAUGHTERS PER MEAN	NO. OF MEANS	NO. OF SIREs	NO. CONTEMPOR- ARIES PER MEAN	SD	
				ACTUAL	PREDICTED ^{1.}
2	41507	377	7.34	1009	1118
4	6484	353	10.12	821	871
6	1868	283	12.34	759	765
8	684	200	15.84	680	696
10	302	136	20.80	669	644
12	144	87	20.94	659	621
14	75	54	26.34	592	589
16	45	34	31.52	604	564

$$1.\text{Predicted SD} = \left[\sigma_w^2 \left[\frac{n_{ij} + 14}{16n_{ij}} + \frac{15}{16m_{ij}} + .0625 (1 - R_j) \right] \right]^{1/2}$$

TABLE 12. DAUGHTER VARIABILITY BY REPEATABILITY (R) CLASS.

R	NO. OF BULLS	VARIANCE
20 to 29	526	1,682,792
30 to 39	196	1,683,871
40 to 49	123	1,620,817
50 to 59	112	1,724,699
60 to 69	111	1,641,467
70 to 79	135	1,550,489
80 to 89	178	1,648,655
90 to 99	193	1,687,091

are presented in Table 13. Considering both negative and positive changes in PD values, there was little evidence of a relationship between the average change from initial to final proof, and the bull's initial R. Bulls of a low R were more likely to change in one direction or the other (as evidenced by the larger standard deviation and range for these bulls), but the probability of a positive change appears to approximately equal that of a negative change. The results are summarized in Table 14. Positive PD changes were greater for higher R bulls and negative PD changes were greater for lower R bulls. When changes in both directions are considered, there was little effect of initial R on the average PD change. Table 15 contains the analysis of PD changes for bulls whose R of the final proof was unrestricted in an attempt to increase numbers of bulls. The conclusions obtained from the earlier data set are reinforced by these data. There was little relationship between change in PD value and sire's initial R.

Use as an indicator of non-randomly overestimated sire proofs was the third measure of a supplemental value for R. The difference between average PD changes for AI bulls and average PD changes for non-AI bulls was defined to be a non-random error of genetic merit estimation. Table 16 presents the PD changes for AI and non-AI bulls whose initial proof had R values ranging from 20% to 40%. AI bulls whose final R was 90% or greater had a larger mean negative change in PD and had a broader range of PD changes than non-AI bulls. Although initial PD was at a lower level and tended to decrease more for AI bulls when compared to non-AI bulls, the small number of bulls in this analysis causes this conclusion to be made with caution. When final R was unrestricted (to increase the

TABLE 13. MEANS, STANDARD DEVIATIONS (SD) AND RANGES OF PD CHANGES
WITH LATE REPEATABILITY ≥ 90

INITIAL R	NO. OF BULLS	MEAN	SD	RANGE	PD DROP (%)
20 to 29	29	-107	289	-903 to 322	62
30 to 39	29	42	225	-345 to 594	38
40 to 49	35	- 37	252	-685 to 445	57
50 to 59	30	25	205	-499 to 572	43
60 to 69	32	61	143	-281 to 349	28
70 to 79	50	46	133	-231 to 435	34
80 to 89	89	20	79	-163 to 218	38
90 to 99	161	7	43	-186 to 191	37
Overall	455	10	121	-903 to 594	40

TABLE 14. REGRESSION OF PD CHANGE ON INITIAL REPEATABILITY¹.

PD CHANGE	REGRESSION COEFFICIENT
Negative	-2.32
Positive	3.05
Pooled	.69

¹. Per 1% change.

TABLE 15. MEANS, STANDARD DEVIATIONS (SD) AND RANGES OF PD CHANGE
WITH LATE REPEATABILITY (R) UNRESTRICTED

INITIAL R	NO. OF BULLS	MEAN	SD	RANGE	PD DROP (%)	LATE R
20 to 29	793	10	143	-903 to 549	48	40
30 to 39	354	20	153	-572 to 676	42	57
40 to 49	238	-3	145	-685 to 445	46	67
50 to 59	171	16	141	-499 to 572	45	76
60 to 69	122	26	115	-281 to 349	39	82
70 to 79	105	22	116	-231 to 435	45	88
80 to 89	142	15	69	-163 to 218	40	91
90 to 99	161	7	43	-186 to 191	37	94
Overall	2086	12	129	-902 to 676	45	61

TABLE 16. PD CHANGES FOR AI AND NON-AI BULLS.

LATE R	TYPE OF BULL	NO. OF BULLS	MEAN	RANGE	INITIAL PD
Restricted ^{1.}	AI	34	-50	-903 to 594	-73
	Non-AI	13	-21	-422 to 259	113
	All	47	-42	-903 to 594	-21
Unrestricted	AI	314	24	-903 to 676	-64
	Non-AI	88	-1	-431 to 336	-14
	All	402	19	-903 to 676	-53

^{1.}Equal to or greater than 90%.

number of bulls), initial PD of non-AI bulls was larger but decreased more. However, the mean final PD was larger for non-AI bulls than for AI bulls. These results indicated little supplemental value for R and suggested that the USDA Contemporary Comparison procedures of estimating genetic merit are more accurate for non-AI proofs than were the Herdmate Comparison procedures.

The value of initial R in predicting final PDM was tested and the results are presented in Table 17. Initial R alone accounted for only 1.3% of the variability of final PDM while initial PDM alone accounted for almost 77% of the variability of final PDM. Thus, initial PDM was a much larger determining factor of final PDM than was initial R. However, final PDM increased approximately 1.53 kg for each 1% increase of initial R. The PDM values of bulls with differing repeatabilities thus could be adjusted for R and bulls could be chosen on adjusted PDM alone. When initial R and PDM were considered together, there was little increase in accuracy of final PDM prediction. When linear and quadratic terms of both initial R and PDM were included in the same model, there was again little increase in accuracy of final PDM prediction over the accuracy of PDM alone. Figure 2 illustrates that initial PDM had a greater effect on final PDM than did initial R. However, at all levels of initial PDM, final PDM was largest when the initial proof had an R of about 70%.

Similar trends are illustrated in Table 18 for prediction of final PD\$. Initial PD\$ alone accounted for almost 75% of the variability of final PD\$ while initial R alone accounted for only 1.2%. However, final PD\$ increased \$.28 for each 1% increase of initial R. Including both

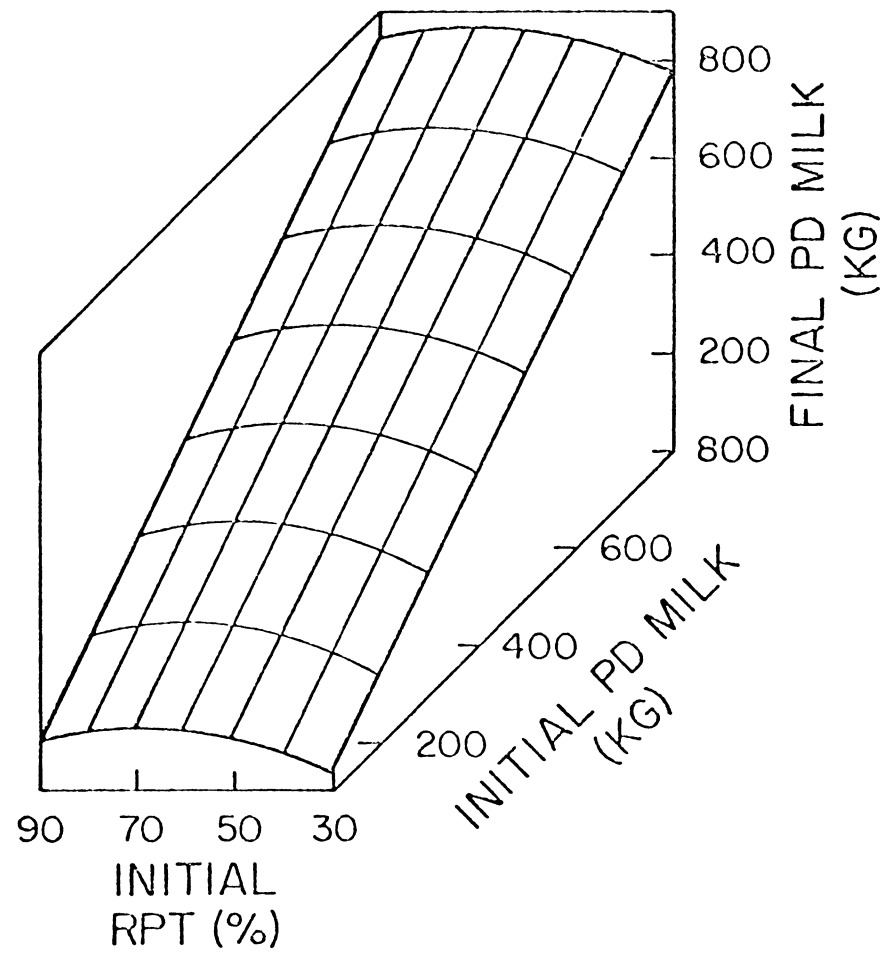


FIGURE 2. FINAL PD MILK PREDICTION FROM INITIAL REPEATABILITY (RPT) AND PD MILK.

TABLE 17. PREDICTION OF FINAL PDM (REPEATABILITY (R) \geq 90).

MODEL	REGRESSION COEFFICIENT	MULTIPLE SQUARED CORRELATION
1. Initial R	1.53092*	.013
2. Initial PDM	1.01986**	.769
3. Initial R	0.67862*	.771
Initial PDM	1.01558**	
4. Initial R ₂	8.69823**	.780
Initial R ²	-0.06314**	
Initial PDM ₂	1.01576**	
Initial PDM ²	0.00003	

*Significant (P<.05)

**Significant (P<.01)

TABLE 18. PREDICTION OF FINAL PD\$ (REPEATABILITY (R) \geq 90).

MODEL	REGRESSION COEFFICIENT	MULTIPLE SQUARED CORRELATION
1. Initial R	0.27952*	.012
2. Initial PD\$	1.01026**	.747
3. Initial R	0.09469	
Initial PDM	1.00659**	.749
4. Initial R ₂	1.79146**	.760
Initial R	-0.01339**	
Initial PD\$ ₂	1.00261**	
Initial PD\$ ²	0.00050	

*Significant (P<.05)

**Significant (P<.01)

initial PD\$ and R in the final PD\$ prediction model resulted in almost no increase in accuracy over the value of initial PD\$ alone to predict late PD\$. Consideration of linear and quadratic terms of both initial R and PD\$, resulted in little increase of accuracy although all terms except PD\$ quadratic were significant (P .01).

Cost of Repeatability

Characteristics of the 215 Holstein bulls that had only production information are presented in Table 19. Several models were examined for explaining semen cost of bulls lacking type evaluations. The model with maximum multiple squared correlation (.31) included PDM, R, PD fat percent, and the interaction between PDM and R (PDM x R) in addition to stud. Standard partial regression coefficients and F values for the semen cost model are presented in Table 20. Repeatability was not a significant factor affecting differences in semen cost after consideration of other factors. Semen cost was most affected by factors peculiar to the particular stud that sold a bull's semen. Bull's PD fat percent significantly affected semen price, but its standard partial regression coefficient was smaller than those for PDM and the interaction between PDM and R. This may be partially due to the fact that little variability in PD fat percent, relative to other factors considered, was observed in the data examined. No other factor in the model had a significant effect on semen cost.

Characteristics of the 94 Holstein bulls having both production and type information are presented in Table 21. Several models were also examined for explaining semen cost for this data set. The model

TABLE 19. CHARACTERISTICS OF 215 HOLSTEIN BULLS WITH ONLY PRODUCTION INFORMATION

FACTOR	MEAN	STANDARD DEVIATION	RANGE
Semen Cost	5.57	4.39	2 to 30
PDM	392	140	104 to 831
PD Fat Percent	-0.07	.08	-0.29 to 0.16
PD\$	72	28	12 to 157
R	66	20	27 to 99

TABLE 20. REGRESSION COEFFICIENTS AND F VALUES FOR SEMEN COST MODEL
 EXAMINING BULLS WITH ONLY PRODUCTION INFORMATION¹.

FACTOR	STANDARD PARTIAL REGRESSION COEFFICIENT	F VALUE
PDM	0.23375	1.07
PD Fat Percent	0.13974	4.50*
R	0.12953	.60
PDM x R	0.33739	.22
Stud	---	5.74**

¹. Multiple squared correlation = .31.

*Significant (P<.05)

**Significant (P<.01)

TABLE 21. CHARACTERISTICS OF 94 HOLSTEIN BULLS WITH BOTH PRODUCTION
AND TYPE INFORMATION.

FACTOR	MEAN	STANDARD DEVIATION	RANGE
Semen Cost	6.17	4.34	2 to 25
PDM	367	168	107 to 874
PDT	-0.11	.81	-2.25 to 1.87
PD Fat Percent	-0.06	0.10	-0.28 to 0.26
PD Dollars	68	33	6 to 162
R For Production	81	19	29 to 99
R For Type	69	21	28 to 99

with maximum multiple squared correlation (.55) included predicted differences for milk (PDM) and type (PDT) in both linear and quadratic forms, repeatability for milk (R) and type (PDTR) evaluations, interactions between PDM and R, PDT and PDTR, and PDM and PDT, in addition to stud. Standard partial regression coefficients and F values for the semen cost model for these bulls are presented in Table 22. When both production and type information were available, R had a highly significant, negative effect on semen cost. A 1% increase in R resulted in \$.16 decrease in semen cost. PDTR also significantly affected semen cost. However, interpretation of these individual effects of R for PDM and PDT may be misleading due to the R by PDM and PDTR by PDT interactions. Factors peculiar to the stud significantly affected semen cost. Increases in PDM and PDT were associated with a non-significant decrease in semen cost. Other factors considered did not significantly affect semen cost. The difficulty of interpretation led to a graphical examination of this model in an attempt to further explain differences in semen cost. For simplicity of illustration, repeatabilities for PDM and PDT were assumed equal.

Prediction of semen cost from various combinations of R and PDM when PDT is held constant at 0.0 is illustrated in Figure 3. At low values of PDM, an increase in R resulted in a semen cost decrease. However, for PDM values greater than about 300 kg, an increase in R resulted in an increase in semen cost. Repeatability had a greater cost when PDM values were higher. When PDT was held constant at +2.0, for any value of PDM from +100 to +900 kg, an increase in R resulted in an increase in semen cost. The positive effect of PDT on semen cost is

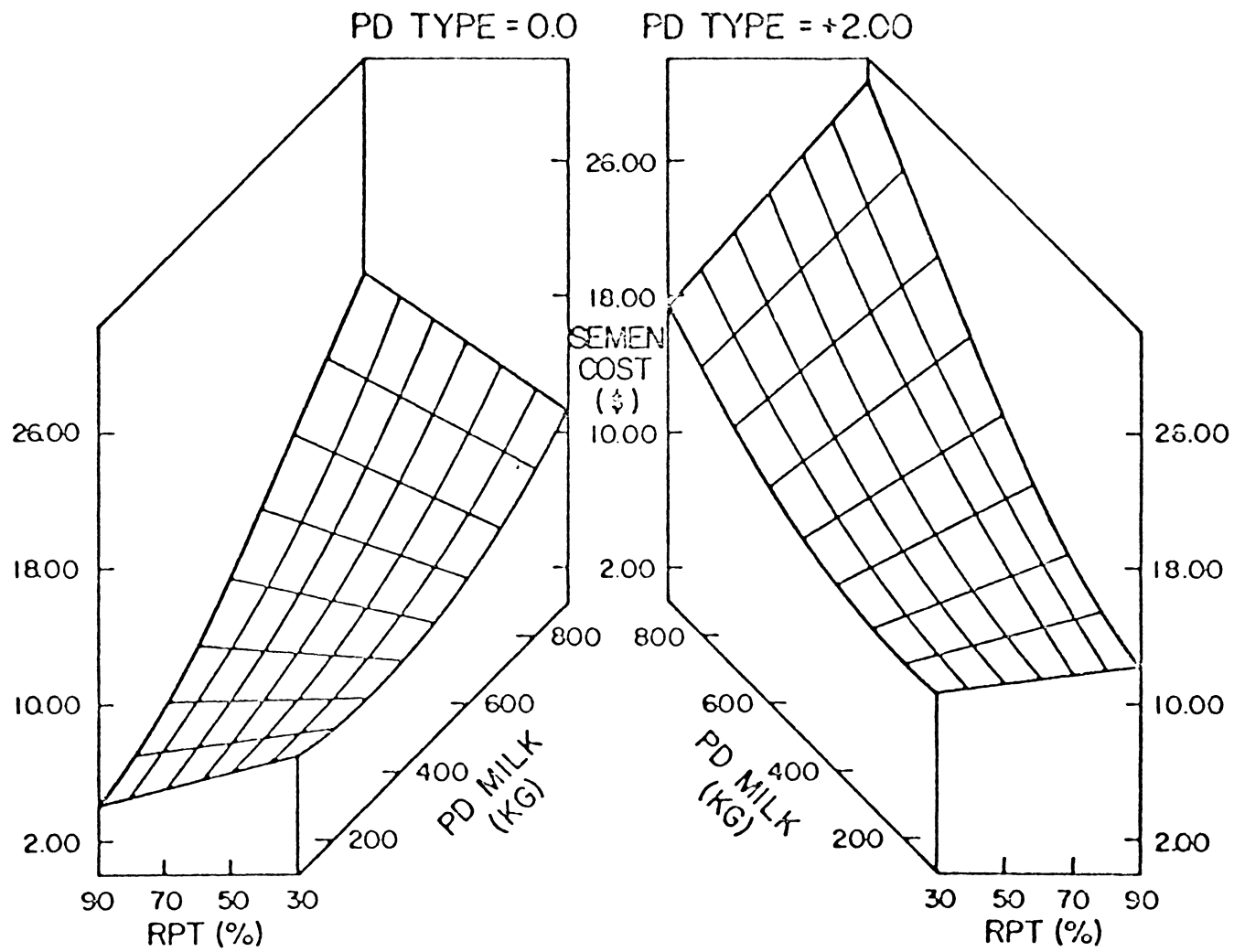


FIGURE 3. SEMEN COST PREDICTION FROM REPEATABILITY (RPT) AND PD MILK WITH PD TYPE CONSTANT.

TABLE 22. REGRESSION COEFFICIENTS AND F VALUES FOR SEMEN COST MODEL
EXAMINING BULLS WITH BOTH PRODUCTION AND TYPE INFORMATION¹.

FACTOR	STANDARD PARTIAL REGRESSION COEFFICIENT	F VALUE
PDM	-0.63131	1.37
PDM ²	0.48861	2.31
PDT	-0.15077	.18
PDT ²	0.12707	2.37
R	-0.70696	6.88**
PDTR	0.39755	7.26**
PDM x R	0.80456	3.23
PDT x PDTR	0.56118	2.94
PDM x PDT	0.15009	.66
Stud	---	3.77*

¹. Multiple squared correlation = .55.

*Significant (P<.05)

**Significant (P<.01)

apparent from comparisons of semen costs at PDT of 0.0 and +2.0. Perhaps the most significant feature of these results was the interaction between measures of a bull's performance (PDM and PDT), and the accuracies of those measurements (R and PDTR), in determining semen cost. For bulls with average performance, greater accuracy resulted in decreased semen cost. In contrast, semen cost increased with greater accuracy as the level of performance increased. Similar results were reported by Bell and McDaniel (1977).

Semen cost prediction from various combinations of PDT and PDM when R is held constant at 30% is illustrated in Figure 4. An increase in PDT from -1.0 to +2.0 was associated with an increase in semen cost at all PDM values for +100 to +900 kg. Semen cost increased at an increasing rate as PDT increased when R was held constant at 90%. Equivalent increases in PDT had a greater effect on semen cost at a higher R. These results again illustrate the importance of interaction between measures of performance and the accuracy of those measures.

An effort was made to combine results pertaining to a supplemental value for R with those indicating a direct cost of R. The equation predicting last PDM (LPD) from initial PDM and initial R suggested a supplemental value for R in terms of a higher expected LPD with higher initial R. Expected LPD values were combined with predicted semen costs to determine semen cost per kg LPD for various combinations of initial PDM, initial R and PDT. While these values are not interpretable in terms of profit or loss, they do represent the supplemental value of R relative to its cost. Values are presented in Figure 5 for PDT of 0.0 and +2.0.

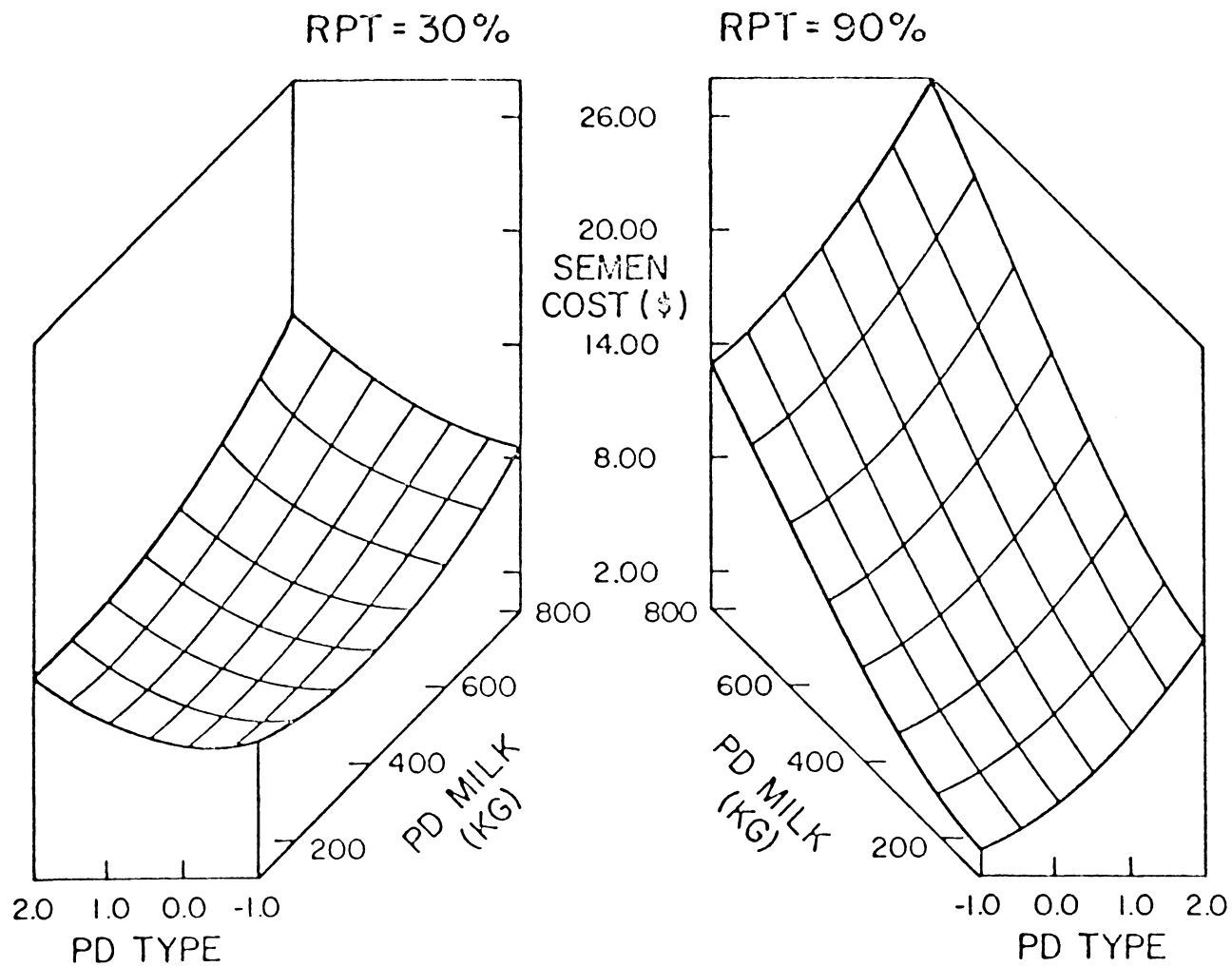


FIGURE 4. SEMEN COST PREDICTION FROM PD TYPE AND PD MILK WITH REPEATABILITY (RPT) CONSTANT.

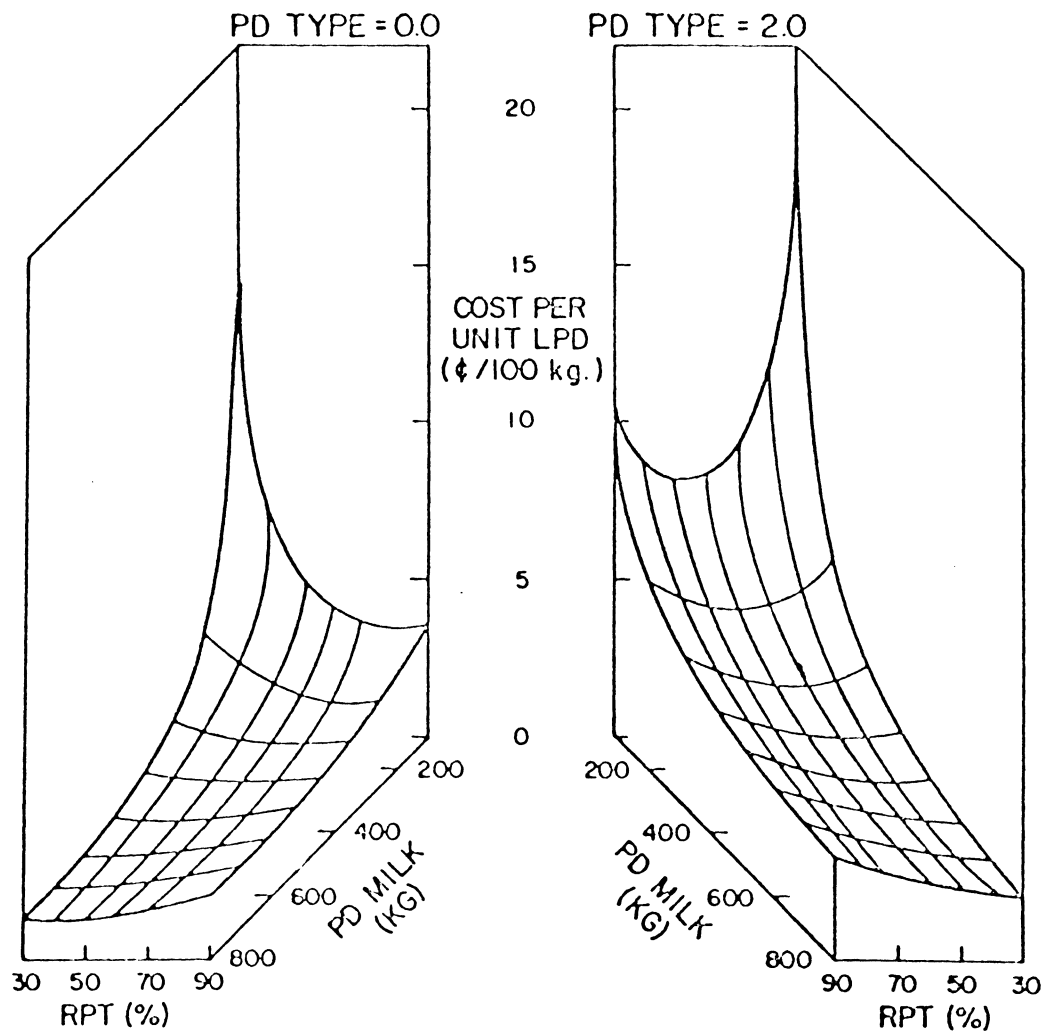


FIGURE 5. SEMEN COST PER UNIT LAST PD MILK (LPD) WITH PD TYPE CONSTANT.

Results in Figure 5 indicate that high PDM bulls represent semen purchases that are far more economical in terms of cost per unit LPD. Comparisons between PDT of 0.0 and +2.0 also indicate a significant cost for type at constant PDM. The relationship between cost per kg LPD and initial R again illustrates the importance of interactions between R, PDM, and PDT. For average PDT (0.0), minimum cost per kg LPD occurs at the lowest R values for PDM exceeding about 500 kg, and at the highest R values for lower PDM. For high PDT (+2.0), minimum cost again occurs at the lowest values of R for PDM exceeding about 500 kg, and at intermediate R for PDM less than 500 kg.

CONCLUSIONS

Although the existence of heterogeneity of sire variances of milk yield was confirmed by this project, little concrete evidence of its causes emerged. The repeatability of sire variances suggest that approximately 25% of the differences in sire variances could be attributed to sire effects. If this value is added to the percentage of variation attributable to specific non-genetic factors which differ among sires (11-18%), the result agrees reasonably well with the total percentage obtained in the sire by herd analysis (40-42%). A logical explanation of a true genetic difference among sires in daughter variability would propose a difference among sires in the number of segregating loci which affect milk yield. However, this explanation is not supported by estimated differences in levels of inbreeding among sires (Henderson, 1976). Discounting heterogeneity of sire variances and concentrating on variance of milk production, the results of this project suggest that sire production measures should not be presented to the dairyman in terms of expected production of future daughters. The mean production of groups of daughters can vary so widely from predicted production that such interpretation could cause erosion of user confidence in the entire sire summary procedure. Variability of milk yield is so large that production genetic merit indicators should be used solely to rank bulls and subsequent selection decisions should be based upon a bull's relative rank.

Analysis of the objective measures of supplemental values for R revealed that R should not receive equal emphasis with genetic merit in

sire selection. PD values of bulls with low R, when compared with high R, were more likely to change, but the probability of a positive change was approximately equal to that of a negative change. Final PDM and PD\$ were both significantly affected by initial R, but less than 2% of the differences in final PDM and PD\$ could be accounted for by initial R. A specific value of 1.3 kg per 1% increase of R may be used to adjust PDM values of bulls with differing R values, to enable selection decisions to be based solely on estimated genetic merit. However, R cannot be considered without giving attention to its interaction with PDM and PDT. Analysis of the cost of final PDM suggested that dairymen selecting bulls of high genetic merit for type and production may make semen purchases that are more economical by choosing sires with low to moderate R. In contrast, for dairymen whose sire selection programs are less intense, the supplemental value of R tends to outweigh its costs such that most economical semen purchases are made at moderate to high R.

BIBLIOGRAPHY

- Allaire, F. R. and C. R. Henderson, 1965. Inbreeding within an artificially bred dairy cattle population. *J. Dairy Sci.*, 48: 1366.
- Anonymous, 1976. Spotlight on sires. *The Red Bloodlines*. November, 1976. page 7.
- Bartlett, M. S., 1937. Some examples of statistical methods of research in agricultural and applied biology. *J. Roy. Stat. Soc. Suppl.*, 4:137.
- Bell, B. R. and B. T. McDaniel, 1976. Effect of predicted differences for milk and type on demand for semen of Holstein bulls. *Proc. 71st Ann. Mtg. Amer. Dairy Sci. Assn. North Carolina State University, Raleigh*.
- Bereskin, Ben and A. E. Freeman, 1965. Genetic and environmental factors in dairy sire evaluation. I. Effects of herds, months and year-seasons on variance among lactation records; repeatability and heritability. *J. Dairy Sci.*, 47:347.
- Burdick, J. M. and Lon D. McGilliard, 1963. Interactions between sires in artificial insemination and management of dairy herds. *J. Dairy Sci.*, 46:452.
- Cassell, B. G., E. A. Losee, H. D. Norman and F. N. Dickinson, 1976. Stability of proofs for Holstein bulls sampled under limited and multiherd conditions. *J. Dairy Sci.*, 59:2095.
- Cassell, B. G. and H. D. Norman. 1977. Modified Contemporary Comparison as a predictor for future daughters. *J. Dairy Sci.*, 60:1152.
- Dairy Herd Improvement Letter. *ARS.*, 50-2, April, 1974.
- Dickinson, F. N., R. L. Powell and H. D. Norman, 1976. An introduction to the USDA-DHIA Modified Contemporary Comparison. *U. S. Dept. Agr. Prod. Res. Rpt.* 165:1-7.
- Gilmore, J. A., B. T. McDaniel and G. A. Carlson, 1976. An estimated utility function in choosing Holstein sires. *Proc. 71st Ann. Mtg. Amer. Dairy Sci. Assn. North Carolina State University, Raleigh*.
- Henderson, C. R., 1953. Estimation of variance and covariance components. *Biometrics*, 9:226.
- Henderson, C. R., 1976. Inverse of a matrix of relationships due to sires and maternal grandsires in an inbred population. *J. Dairy Sci.*, 59:1585.
- Johnson, L. E., 1945. The importance of uniformity of get in evaluating a sire. *J. Dairy Sci.*, 28:109.

- Keown, J. F. and Ray H. Kliever, 1976. Do the mates of AI sires influence their proofs? *Hoard's Dairyman*, 121:984.
- Lee, A. J. and C. R. Henderson, 1964. Herd variation in repeatability. *J. Animal Sci.*, 23:851.
- Mao, Ivan, 1976. Bull's daughters vary. *Hoard's Dairyman*, 121:891.
- McDaniel, B. T., F. R. Allaire, and F. N. Dickinson, 1968. The probability for determining which of two bulls has a superior genetic merit. *Dairy Herd Improvement Letter. ARS.*, 44-6.
- McDaniel, B. T., H. D. Norman and F. N. Dickinson, 1974. Variation in genetic merit of sires of herdmates of first lactation cows. *J. Dairy Sci.*, 57:1234.
- McKittrick, J. W., 1976. Back to school on how to use production proofs. *Hoard's Dairyman*, 121:720.
- Miller, R. H. and E. L. Corley, 1965. Usefulness of information on mates of sires in artificial insemination. *J. Dairy Sci.*, 48:580.
- Norman, H. D., B. T. McDaniel and F. N. Dickinson, 1972. Regression of daughter and herdmate milk yield on genetic value of the herdmate's sires. *J. Dairy Sci.*, 55:1735.
- Norman, H. D., R. L. Powell, and F. N. Dickinson, 1976. Modified Contemporary and Herdmate Comparisons in sire summary. *J. Dairy Sci.*, 59:2155.
- Palmer, M. C. and I. L. Mao, 1977. Factors affecting Holstein sire semen sales in Michigan. *Proc. 71st Ann. Mtg. Amer. Dairy Sci. Assn. North Carolina State University, Raleigh.*
- Plowman, R. D. and B. T. McDaniel, 1968. Changes in USDA sire summary procedures. *J. Dairy Sci.*, 51:306.
- Rothschild, M. F., C. R. Henderson, and R. W. Everett, 1976. Bias of sire evaluations for natural service sires and age of sire. *J. Dairy Sci.*, 59:2091.
- Van Vleck, L. D. and G. R. Barr, 1963. Genotype and variability in dairy lactation records. *Genetics*, 48:1441.
- Van Vleck, L. D., 1968. Variation of milk records within paternal-sib groups. *J. Dairy Sci.*, 51:1465.
- Wadell, L. H., L. D. Van Vleck, and C. R. Henderson, 1960. Distributions and parameters of records of AI daughters. *J. Dairy Sci.*, 43:878.

Wright, E. E., 1975. When choosing bulls to use in your breeding program. Hoard's Dairyman. 120:991.

**The vita has been removed from
the scanned document**

THE IMPACT OF SIRE DAUGHTER VARIABILITY
AND REPEATABILITY ON THE ACCURACY
OF SIRE SELECTION

by

John Springmann Clay

(ABSTRACT)

Modified contemporary sire evaluation procedures, using first lactation records only, were used to compute Predicted Differences for all Holstein bulls having 10 or more daughters calving in any calendar year between January 1964 and June 1973. Significant heterogeneity of within sire daughter milk yield variance was found although only 18% of the differences in sire variances could be accounted for by factors examined. Repeatability of sire daughter variances suggested that 23 to 32% of the differences in sire variances could be attributed to sire differences. Heritability estimates of sire daughter variance were low to moderate and ranged from $-.14$ to $.29$. The standard deviations of mean milk yield, adjusted for number of contemporaries, of paternal half-sibs in the same herd-year-season ranged from 999 to 522 kg for groups of 2 to 16 daughters.

Repeatabilities of contemporary comparison sire summaries were examined to determine their impact on the accuracy of sire selection. A set of 455 bulls have final Repeatabilities exceeding 90% was used to examine the effects of initial Predicted Difference and Repeatability on final Predicted Difference. The effect of initial Predicted Difference on final proof was highly significant ($P .01$) and was the most important

predictor of final proof. However, the linear effect of Repeatability on final proof was also significant ($P < .05$), indicating an increase of 1.53 kg in final proof per 1% increase in initial Repeatability. Relationships between semen cost and Repeatability indicated that low Repeatability bulls frequently represented most economical semen purchases.